Saridis, G. M., Peng, S., Yan, Y., Aguado, A., Guo, B., Arslan, M., Jackson, C., Miao, W., Calabretta, N., Agraz, F., Spadaro, S., Bernini, G., Ciulli, N., Zervas, G., Nejabati, R., & Simeonidou, D. (2016). LIGHTNESS: A Function-Virtualizable Software Defined Data Center Network with All-Optical Circuit/Packet Switching. *Journal of Lightwave Technology*, *34*(7), 1618-1627. [7359113]. https://doi.org/10.1109/JLT.2015.2509476

Peer reviewed version

Link to published version (if available):
10.1109/JLT.2015.2509476

LIGHTNESS: A Function-Virtualizable Software Defined Data Center Network with All-Optical Circuit/Packet Switching

George M. Saridis, *Student Member, IEEE*, Shuping Peng, Yan Yan, Alejandro Aguado, Bingli Guo, Murat Arslan, Chris Jackson, Wang Miao, Nicola Calabretta, Fernando Agraz, Salvatore Spadaro, Giacomo Bernini, Nicola Ciulli, Georgios Zervas, *Member, IEEE*, Reza Nejabati, *Member, IEEE*, Dimitra Simeonidou, *Member, IEEE*

***Abstract*— Modern high-performance Data Centers are responsible for delivering a huge variety of cloud applications to the end-users, which are increasingly pushing the limits of currently deployed computing and network infrastructure. All-optical dynamic data center network (DCN) architectures are strong candidates to overcome those adversities, especially when they are combined with an intelligent software defined control plane. In this paper, we report the first harmonious integration of an optical flexible hardware framework operated by an agile software and virtualization platform. The LIGHTNESS deeply-programmable all-optical circuit and packet switched data plane is able to perform unicast/multicast switch-over on-demand, while the powerful Software Defined Networking (SDN) control plane enables the virtualization of computing and network resources creating a virtual data center (VDC) and virtual network functions (VNF) on top of the data plane. We experimentally demonstrate realistic intra data center networking with deterministic latencies for both unicast and multicast, showcasing monitoring and database migration scenarios each of which is enabled by an associated network function virtualization (NFV) element. Results demonstrate a fully-functional complete unification of advanced optical data plane with an SDN control plane, promising more efficient management of the next-generation data center compute and network resources.**

***Index Terms*— Data Center Networking, Multicast, Network Function Virtualization, Optical Circuit Switching, Optical Packet Switching, Software Defined Networking, Virtual Data Center, Virtual Network Function**

I. Introduction

Conventional internet and telecom data centers are facing the rapid development of a wide range of emerging services and applications, such as 4G/5G, Internet of Things (IoT), High Definition (HD) multimedia streaming, multi-tenancy, Big Data management, cloud storage and processing, etc. Since the demand for these online web-based services is escalating tremendously, Data Center Networks (DCN) will have to accommodate increasing amounts of traffic. The vast majority of this storage and data exchange traffic, contrary to popular belief, does not run between the data center itself and the end-users, but usually resides within the data center; between servers of the same rack and within different racks and clusters [2]. Legacy multi-tier DCN architectures (fat-tree, etc.) are unable to provide the required network efficiency, flexibility, programmability and topology plus wiring low complexity. Next generation data center infrastructure is expected to support more advanced IT facilities, increased storage and processing capabilities along with high network bandwidth, more efficient utilization of network and computing resources, lower interconnection latency values and finally reduced power consumption and operational expense (OPEX) [3], [4].

Advanced photonic technologies have great potential to meet the above network capacity, latency and energy efficiency requirements, and in conjunction with high-speed electronics on the edge of the network, they could constitute reliable intra-DCN communication hardware solutions. Hybrid [5], [6] and all-optical DCN designs have been proposed recently. Some of latter utilize WDM with AWG-based interconnects [7], or offer WDM/TDM interconnection with spectrum selective switch (SSS) -based Top of the Rack/Cluster (ToR/ToC) switches [8], [9]. Others utilize SDM/TDM with multi-element fibers and PLZT-based fast switches for intra-DCN communication [10], or scalable architecture using nanoseconds optical packet switches [11]. The above all-optical architectures claim dynamic flexible bandwidth, increased capacity and ultra-low latency interconnection, outperforming the capabilities of

Manuscript received xx xx, 2015; revised xx xx, 2015; accepted xx xx, 2015. Current version published xx xx, 2016. This work was presented in part as a post deadline paper at ECOC 2015 [1].

This work is supported by EC FP7 grant no. 318606, LIGHTNESS project and partially by EPSRC EP/I01196X: The Photonics Hyperhighway.

G. M. Saridis, S. Peng, Y. Yan, A. Aguado, B. Guo, M. Arslan, C. Jackson, G. Zervas, R. Nejabati and D. Simeonidou are with the High Performance Networks, University of Bristol, Bristol, United Kingdom (e-mail: George.Saridis@bristol.ac.uk).

W. Miao, N. Calabretta are with COBRA, Eindhoven University of Technology, Eindhoven, Netherlands. (e-mail: w.miao@tue.nl).

F. Agraz, S. Spadaro are with the Universitat Politecnica de Catalunya, Barcelona, Spain (e-mail: agraz@tsc.upc.edu).

G. Bernini, N. Ciulli are with the Nextworks, Pisa, Italy (e-mail: g.bernini@nextworks.it).

current intra-DCN designs based in commercially available electronic equipment.

Currently the management of the various DCN control functions is often manual and static. It must move towards to more dynamic and automated solutions in order to provide higher availability and adaptable provisioning of the available network resources. The trend of future data centers is moving towards the adoption of control and management approaches based on SDN solutions combined with resource virtualization and distributed cloud computing [12]. SDN-enabled optical network devices, provisioned through protocols like OpenFlow (OF) [13], along with the virtualization of fundamental DCN functions (such as database migration, multicasting, monitoring, etc.), can lead to multiple abstraction layers of the DC physical resources. This would offer the DC operator a handy toolset and a full view of the system while facilitating a more efficient management of the infrastructure [14].

In this paper, we present and experimentally demonstrate, for the first time, a fully SDN-programmable intra-DCN architecture including network function virtualization capabilities for even more effective network control than what is currently deployed. It is the first demonstration of NFV and SDN functionalities (such as monitoring and database migration) entirely integrated with an advanced all-optical physical layer. Our data plane is capable of performing Optical Circuit Switching (OCS) to Optical Packet Switching (OPS) switch-over and vice versa on-demand, providing intra-DCN connectivity with low deterministic latency and variable bandwidth granularity. LIGHTNESS system supports the construction and reconfiguration of multiple VDCs, respecting their main requirements like isolation, resource orchestration and allocation, etc. VDC are pools of virtual compute, memory, storage and virtual network resources abstracted from the physical layer. Initiated by those VDCs and triggered by the control plane, the OCS/OPS switch-overs operate either unicasting or multicasting functions, according to the NFV requirements.

II. Overall Architecture

The introduced architecture, shown in Fig. 1, represents a next-generation ultra-programmable data center network based on both optical circuit and optical packet switching technologies. Having a closer look of the DCN design, server blades within the same racks are interconnected via novel optical Network Interface Cards (NIC) and all-optical ToR switches to the rest of the reconfigurable DCN. The FPGA-based NICs employ SDN-enabled hybrid OCS/OPS interfaces that support programmable composition and transmission of optical packets with correlated labels or Ethernet frames [15]. In order to support direct OCS multicasting capabilities inside each rack, optical power splitters are also attached on each of the optical ToRs.

We propose a high-radix space switch as an optical ToR, because in conjunction with the advanced NICs on each server, they are able to eliminate the need for electronic ToR switches. This scheme offers lower interconnection latency and potential reduction in power consumption of the overall network, due to absence of frequent O/E/O conversions. It also supports full bandwidth transparency, since optical switches are totally agnostic of the link bitrate, the network protocol or the modulation format that is being used.

Within each cluster, the optical ToRs are connected to a top of the cluster (ToC) switch, as shown in Fig. 1. The ToC consists of a flexible optical network including a high-radix optical switch, serving also as an optical backplane, optical power splitters, a wavelength/spectrum selective switch (WSS/SSS) and a 4×4 optical packet switch. The optical power splitter at that point of the network realizes OCS multicasting among different servers of different racks within and between clusters. The WSS enables grooming/dividing multiple inter-cluster communication channels and traffic in an elastic manner. A passive filtering device, such as an AWG, could perform similar tasks and be currently used instead of the WSS. However, we prefer using a WSS due to its increased

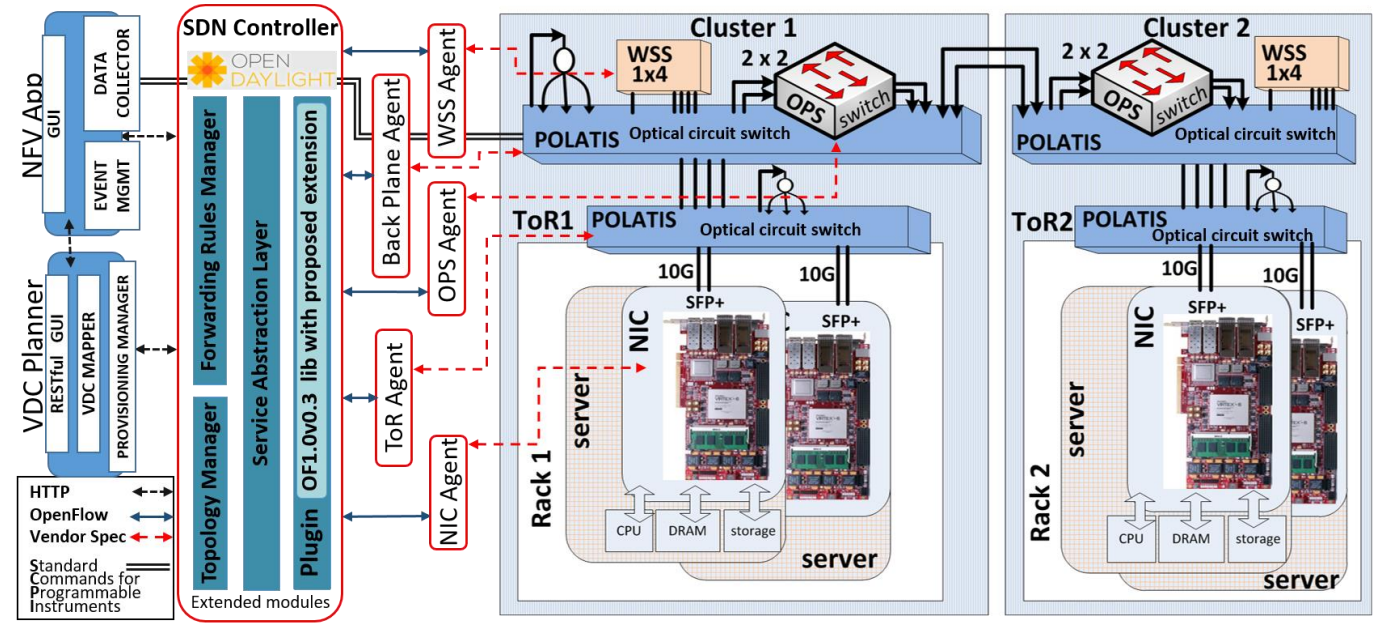

Fig. 1. Overall Data Center Network architecture; experimental data plane (center-right), control plane (left) and virtualization schemes (far left).

programmability, flexibility and bandwidth variability, while a WSS/SSS could easily adapt on future possible standards with more spectral efficient and narrower channel bandwidths than the present ones of 50GHz and 100GHz.

The optical backplane also includes SDN-enabled OPS nodes which can perform nanosecond-fast packet switching, multicasting as well as supporting of monitoring capabilities of optical packet reception/contention. To achieve nanoseconds forwarding operation, the OPS processes the optical packets according to the optical label provided by the NIC [16], while the SDN controller has the role to provision the look-up tables of the OPS and NIC. This allows de-coupling the fast (nanoseconds) forwarding operation of the optical data plane to provide time domain fast statistical multiplexing capability to the DCN, from the slower SDN control plane and VDC planner application for the virtualization of the DCN. Taking advantage of statistical multiplexing, the OPS can also offer efficient and flexible bandwidth utilization therefore lowering the required number of optical ports at the backplane to guarantee the required connectivity. In combination with the nanoseconds label detection and switching control, bursty traffic demands are better served with higher degree of bandwidth granularity, low-latency and versatile per-packet processing intelligence.

Each of the NICs, optical ToRs, optical back plane, OPS switching nodes and WSS switches are completely controllable by the logically centralized SDN controller through a uniform control software interface, exposed by a dedicated device agent, as shown in Fig. 1. The agents gain useful information from the hardware devices, keeping the SDN controller updated with the current state of the network, while they also push the control commands to the physical layer devices. Furthermore, on top of the SDN controller, the VDC planner and the NFV applications offer one more layer of abstraction and virtualization of the deployed infrastructure.

In summary, the programmable data plane enables the SDN-based DCN control plane to build and reconfigure the physical layer topology, by dynamically provisioning appropriate cross-connections in the optical backplane to match the different applications' requirements. In addition, based on the DCN requirements and data flows, the FPGA-based hybrid OCS/OPS NIC can be configured by the SDN controller on-demand along with the optical ToRs, ToCs and OPS switches, thus achieving unicast and/or multicast communication among servers.

III. All-Optical Reconfigurable Data Plane

The data plane test bed used for the experiments consists of four rack-mounted Dell PowerEdge T630 servers each equipped with an advanced FPGA-based NIC board utilizing 10G SFP+ transceivers, serving as the programmable interface of the computer blades to the optical network [15]. On those servers various virtual machines (VM) are able to reside, one of which also hosts the SDN-controller. All servers are connected to the 192×192 port Polatis optical circuit switch, which acts as ToR and as the optical backplane on top of each cluster. Polatis beam-steering backplane switch inserts around 1 dB of loss per cross-connection (OXC), so multiple OXCs and hops are achievable without major power and signal quality penalties. As mentioned in the overall architecture, we utilize a 1×4 optical power splitter, two 1×4 WSS and one SOA-based 4×4 OPS [16]. All the above equipment is attached to the optical backplane on top of the cluster.

The functionality of the novel NIC includes network interface functions, programmable aggregation and segregation functions, OCS/OPS switching and layer 2 switching functions. The FPGA-based hybrid OCS/OPS NIC using NETFPGA SUME development board has been designed to plug directly into a server, and replace the traditional NIC. In the prototype design, it has an 8-lane Gen3 PCIe interface for DRAM communication, one 10 Gb/s interface for getting commands from SDN control agent and sending feedback, two OCS/OPS hybrid 10 Gb/s SFP+ ports for inter-server communication and an OPS label pin interface connected to the OPS label generator. The SFP+ transceivers' channels are in the 1550nm region and ITU grid-spaced, in order to be compatible with the LCoS-based WSS and SOA-based OPS switches, which both normally operate in that frequency band.

The 1×4 optical power splitter is used to accomplish OCS one-to-four multicasting scenarios. The WSSs are used for grooming inter-cluster traffic carried by channels from different servers or racks into an inter-cluster WDM super-channel. In the destination cluster, the local WSS de-multiplexes the super-channel and switches the channels to the receiving racks and servers.

The OPS switch is able to rapidly switch optical packets with a reconfiguration time of 20 nsec. The OPS is based on a modular WDM architecture that allows scalability of the number of ports beyond 128×128 while the highly distributed control and parallel packet processing allows port count independent reconfiguration time. A fully equipped 4×4 prototype including optical label processing, optical switching fabric and controller has been realized in the LIGHTNESS framework. The modular architecture allows the 4×4 OPS prototype to logically perform as two 2×2 OPS. This is also how we use it for our studies, one 2×2 OPS node per cluster, as shown in the overall setup in Fig. 1. The electrical label bits generated by each NIC, are encoded in an in-band optical RF tone label [16] by a prototyped label generator. The in-band optical labels are then coupled to each of the optical packets. At the OPS node, the optical label of each packet is filtered out, processed and matched with the look-up table by the switch controller in order to determine the packets destination. As shown in the time traces of Fig. 2, depending on the combination of the values of the labels, different (or no) outputs are activated for each 38.4 μsec timeslot. Multicasting is enabled when two label bits have been set as "11".

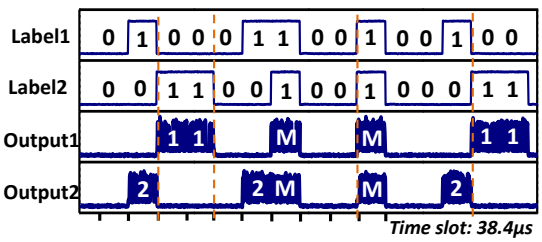

Fig. 2. Time traces of labels and OPS switch outputs for normal switching operation and multicasting

IV. SDN-enabled Control Plane

For this experiment, OpenDaylight (ODL) is used as the SDN controller, and OF agents for Polatis, WSS, OPS switch and hybrid OCS/OPS NIC were developed to enable SDN-based programmability, as shown in Fig. 1. In addition to the OpenFlow extensions, as previously reported in [17], the NIC OF agent is further extended to allow configuration of the duration of the generated OPS packets.

Also, ODL internal software modules are extended to support some other network device specific features. For example, regarding the OPS and WSS ports, the switch manager and Service Abstract Layer (SAL) were extended to record the supported wavelength and supported spectrum range respectively, both of which are used to validate the configuration. Furthermore, the transmitted optical packet statistics can be collected and maintained by the statistics manager. In order to properly configure the above optical devices, the forwarding rules manager has been extended to construct the required set of configuration information e.g. label & output for the OPS switch; central frequency, bandwidth & output for the WSS and match, label and output for the NIC (which is optional for OPS).

The FPGA-based hybrid OCS/OPS NIC communicates with the OF agent through a bidirectional 10Gbps SFP+ Ethernet interface. The commands and information are encapsulated in a 1504 Byte Ethernet Frame (VLAN). Furthermore, through the extended ODL, various applications can communicate directly with the hardware using the RESTful interface.

V. Virtual Data Center (VDC) Management And Applications

In this paper the relation between our approach and the ETSI NFV proposed architecture [18] is not directly mapped because a) we are handling a single type of function, not multiple VNFs and b) this function is also lightweight and does not have many requirements in terms of infrastructure (computation or storage). For this reason, the multiple monitoring functions run as multiple threads that expose information through a RESTful API interface. In order to relate our application with ETSI NFV solution, we can assume that our application's core acts as a NFV orchestrator, which instantiates new individual threads for each VDC user by request. This NFV orchestrator (core) runs an interface (RESTful) on a fixed IP address; thus it handles user requests by checking HTTP authentication and then allocates a specific URL within the same IP to each user. In our experimental demonstration this was simplified, since we were running a single all-optical VDC at a time. VNF management functionalities are handled by this core: instantiation, update (in terms of port monitoring, changes on the VDC, optical cross-connection handling, etc), scaling (not applicable) and termination (same lifecycle termination as VDC). Virtualized Infrastructure Managers are also avoided in this experiment, since, as mentioned above, our instances are running as threads within one single virtual machine, and not as individual VMs or containers.

For the purposes of the experiment, two control plane applications have been implemented and deployed on top of the ODL: a virtual data center planner (VDC Planner) and a virtual network monitoring function (monitoring VNF). The VDC planner allows to compose and provision virtual network slices within the DCN enabling thus multi-tenant data centers. It consists of a graphical user interface (GUI) developed in HTML/JavaScript that interfaces with a backend application developed in Python 2.7 able to interact with the ODL controller. The user can access to the GUI with any existing browser, and create dynamically a VDC request, which is shown in a graph and a table. The parameters that the user can specify are: servers to be used, links to be created, technology for each link and multicast properties for servers and links. Other parameters that are available (not mandatorily applied for the algorithm) are: required bandwidth and bi-directionality of a given link. The application receives a set of requirements for the VDC and generates a bunch of static flows to be pushed in the DCN by ODL, distributed among the different technologies (NIC cards, OPS Switches and OCS backplane). This set of flows is generated in JavaScript object notation (JSON) format

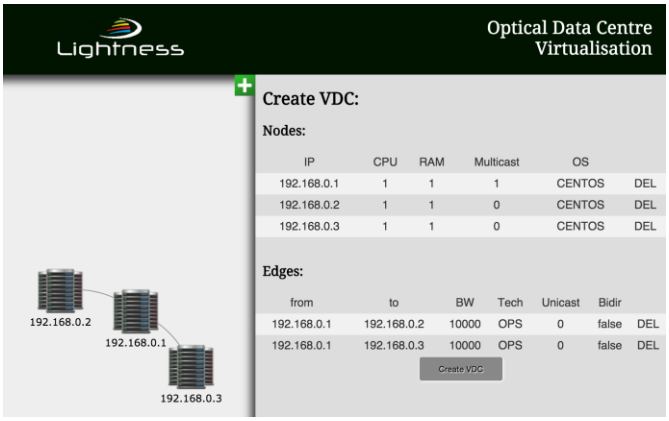

Fig. 3. VDC planner application with OPS or OCS multicasting options

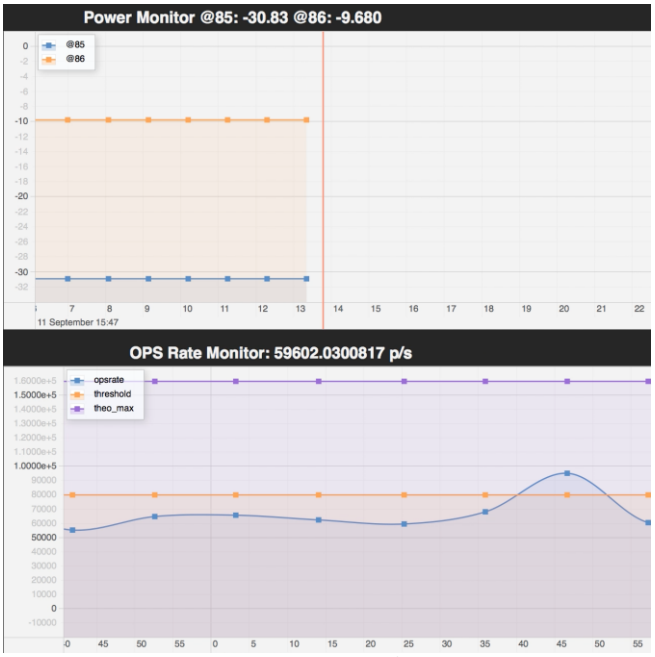

Fig. 4. Monitoring VNF. Power monitoring of the Polatis ports associated to the output of the OPS switch (top) and monitoring of the current packet rate, threshold and theoretical maximum of the OPS node (bottom)

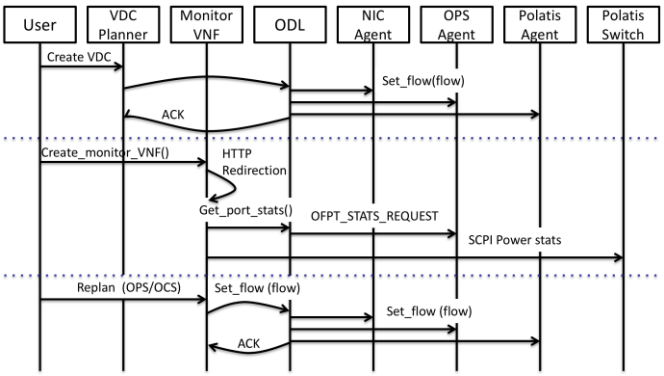

Fig. 5. VDC creation, monitoring function initiation and replanning workflows

and sent to the ODL controller via a RESTful interface. Fig. 3 shows an example of a VDC request using our aforementioned VDC planner. The user specifies for this use case three hosts, one of them selected as a multicast node, which will send the content to the other two. At the same time the request contains OPS technology as multicast solution for the VDC.

Furthermore, in our proposed LIGHTNESS architecture, data center performance can be evaluated in two different ways, by either monitoring physical network impairments or application impairments. For our all-optical data center solution, we expose two different optical network indicators to the user (optical power and packet rate monitoring), which can directly affect user's application performance. Application-related information could also be retrieved with only a software extension to the FPGA-based NICs, but it was out of the scope of this work.

Thus, when a tenant's VDC has been deployed (in our use cases, a multicast VDC using OPS resources), the user can request a dynamic VNF to monitor various parameters in their DC. The request is a basic HTTP GET request, made through a standard web browser, which is handled by our control & management server creating the monitoring function and redirecting the users web browser using a standard HTTP redirection response (301 Moved Permanently message, with 'Location' header parameter). The user's monitoring VNF starts retrieving network information by means of two different interfaces. RESTful northbound of ODL controller retrieves information of the OPS packet counting and generating a graph of the OPS packet rate on the one end. Standard Commands for Programmable Instruments (SCPI) direct interface is used to make the Polatis OCS switch to retrieve information of the multicast ports used in the test (output of the optical splitter and the OPS switch). This information is plotted in a web-page, showing two graphs to the user with the optical power received in two ports (in dBm) in one of them and the other one of the packet rate in the OPS switch, with a theoretical maximum and a configurable threshold. Two buttons in the bottom part of the monitor, allow the user to start two different workflows: (i) switch the multicast traffic to a second OPS switch by making a backup copy of the content between two servers inside the same rack when the optical power detected drops below an expected value, and (ii) switch the multicast traffic using the optical splitter (OCS multicasting) whenever the OPS packet rate exceeds the threshold, meaning not sufficient OPS resources to cope with the desired VDC service.

Up to this point, the tenant has deployed his own VDC with multicasting capabilities through the OPS switch to provide services to users. At the same time, the tenant has requested his own virtual monitor within his VDC, getting information with different and reconfigurable polling timers (one second for power monitoring, twenty seconds for OPS received packets to avoid time mismatches among OPS agent/controller/ monitoring function). The monitoring VNF allows the VDC tenant to take two different recovery choices based on the monitored data:

(i) The tenant, whose service is experiencing issues in terms of data loss caused by unexpected low power problems, decides that the packet rate within the OPS switch is sufficient to cope with his service. The OPS switch in the VDC (or the involved network up to it) should be avoided to solve any power issue, so the user needs to start a recovery workflow to switch the service to other intra-rack server switching the previous VDC, to a new network using a second OPS switch.

(ii) Based on a high packet rate per second shown by the monitoring VNF, the user decides that, before experiencing any packet loss within his VDC, a reconfiguration of the VDC using

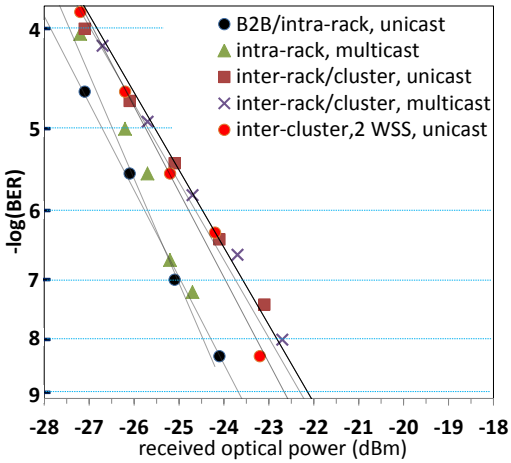

Fig. 6. OCS BER curves for intra/inter-rack unicast/multicast and inter-cluster through 2 WSS with 10GbE traffic.

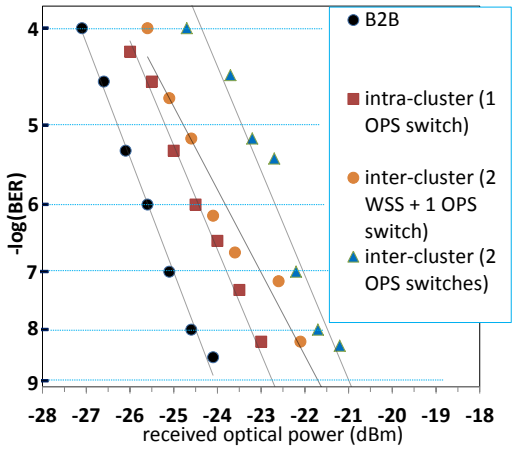

Fig. 7. OPS BER curves for intra-cluster and inter-cluster through WSS and 1 or 2 OPS switches with 10GbE traffic.

a pure OCS network is required. The tenant can use both, the VDC planner and the monitoring VNF (which are connected), to reconfigure the network by moving the content of the first server to a backup one in the same rack and establishing the new set of connections through the optical splitter to maintain the multicast capabilities. Fig. 4 shows a screenshot of the VDC monitoring function, displaying optical power from two ports (top) and current OPS packet rate and the threshold and maximum theoretical OPS packet rate (bottom). The workflows previously explained for the all-optical VDC creation, monitoring function instantiation and for the VDC replanning are shown (from top to bottom, respectively) in Fig. 5.

VI. Experimental Demonstration & Results

For the experimental demonstrator, we combined all the available data plane and control plane resources, as presented in sections III and IV, and validated several intra-DCN interconnection scenarios based on VDC applications' and NFV functions' requests.

First of all, we evaluated the DCN physical layer for intra-rack, inter-rack and inter-cluster unicast and multicast communication by measuring BER for both OCS and OPS switching technologies using real traffic with scrambled PRBS payload from our traffic analyzer, as shown in Fig. 6 and 7. The traffic analyzer feeds the FPGA-based NIC with 10 Gb/s Ethernet traffic, and then the NIC pushes the data to one of its hybrid OCS/OPS ports. When OPS mode is chosen, the NIC, depending on the configuration received from the SDN controller, sets up the optical packet duration, encapsulates certain number of Ethernet frames and releases the optical packet while the label is generated and combined in parallel. Intra-rack communication is realized by going from transmitting to receiving server through the optical ToR for unicast, and through an optical splitter in multicast operation. Inter-rack and inter-cluster are similarly realized by going through multiple Polatis OXCs and/or optical power splitters, while for inter-cluster multiplexed interconnection signals propagate additionally through two WSS for WDM mux/demux and switching purposes. Minor penalties of <2 dB are observed for all OCS interconnection scenarios, as seen in BER curves of Fig. 6. OPS BER plots in Fig. 7 show 1 and <3dB penalties when passing through one (for intra-cluster) and two (for inter-cluster) switches, respectively.

In addition to BER testing of the physical links, we collect network Layer 2 results regarding the interconnection latency from one NIC's DMA to the destination NIC's DMA, excluding the DMA driver's actual delays, which is separately measured for different DMA lengths. Moreover, interconnection throughput is monitored and plotted, exhibiting OCS-to-OPS switch-over and vice versa.

We measure DMA-to-DMA access latency using the traffic analyzer and Ethernet traffic with PRBS payload. The traffic analyzer firstly feeds the transmitting FPGA-based NIC with traffic. Then, NIC pushes the traffic to the all-optical network, using either OCS or OPS, towards the destination NIC. Finally, the latter NIC forwards the received traffic back to the traffic analyzer. We calculate the overall chip-to-chip latency by subtracting the traffic analyzer-to-NIC (and vice versa) delays. Our test and measurements are based on the best possible latency with maximum bitrate, so, for OPS with switching, bitrate is around 3 Gb/s, and for OCS it is around 8 Gb/s. All measured latency values include FPGA physical and logic delays, which can vary depending on the frame length, chosen transmission/switching scheme (OCS or OPS) and FPGA design. Fig. 8 shows unicast and multicast OCS access latencies for all the studied interconnection scenarios; whereas Fig. 9 shows intra/inter-cluster OPS access latencies with and without switching.

When no switching is performed, the clock of the receiving end of the transceiver is continuous, so there is no need for recovering it with extra payload (i.e. preamble dummy key characters). For OPS with switching though, optical packets are formed, a procedure which inserts significant delays due to segregation, aggregation, buffering and clock recovery with extra payload. For instance, when transmitting/receiving 3 packets, the latency is tripled, 33.2 μsec × 3 equal to 99.6 μsec. In the FPGA, the first buffer of the segregation-aggregation part uses numerous FIFOs and store-and-forward techniques, which means it takes 25.6 μsec more time when the OPS with switching mode is chosen. The same happens at the last buffer of aggregation inside the FGPA design, so another 25.6 μsec is added. Since we use two FPGAs (one Tx plus one Rx), the whole segregation-aggregation procedure takes place twice. Thus, in total, the delay is doubled, adding a further 102.4 μsec. Finally, roughly 99.6 + 102.4 μsec equal to a total of 202 μsec of latency, as shown in Fig. 9 (OPS with switching). The receiver of the FPGA-based NIC needs to recover the clock from the receiving traffic if this was lost during the operation. The time of this recovering depends mostly on the network conditions and signal performance. In the case of OPS switching, it needs 25.6 μsec to recover the clock before each

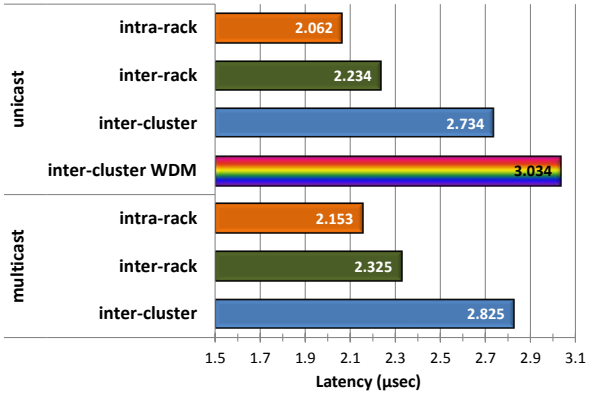

Fig. 8. DMA-to-DMA OCS latency for various intra-DC scenarios

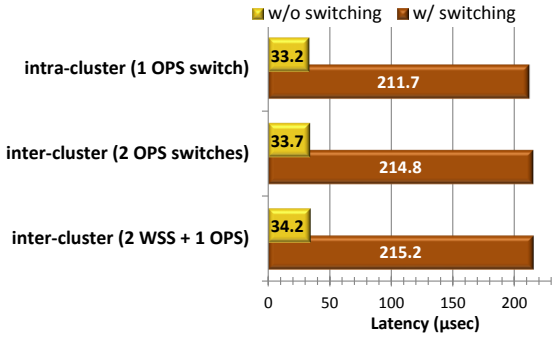

Fig. 9. DMA-to-DMA OPS latency for various intra-DC scenarios

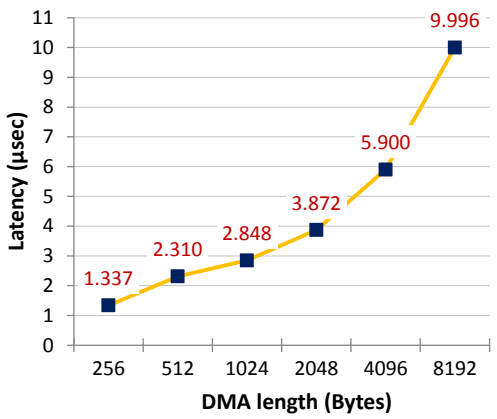

Fig. 10. DMA write and ready latency VS DMA length measured with 256bytes Ethernet packet frame length

packet. Therefore by employing dedicated hardware, such as clock & data recovery circuitry [19] and packet fragmentation ASICs [20], much shorter packet size is sufficient to achieve higher throughput and the latency of the OPS switching can be dramatically reduced below sub-microsecond values (less than 150 nsec have been measured for burst operation).

Additionally, the DMA driver's actual access latency was also measured, as shown in Fig. 10, for different DMA lengths ranging from 256 to 8192 Bytes. It is the time interval we measured when we were pushing Ethernet frames into the DMA in order to be written and then read. We found out that changing the DMA length and keeping fixed 256 Bytes Ethernet frame length, it has an meaningful impact on the total DMA write + read delays. We also discovered that by using larger Ethernet packet frames (e.g. 1024 Bytes), DMA overall access latency critically drops. Last but not least, DMA latency does not rely upon the network state, but it mostly depends on the load and traffic handling policies (interrupts, etc.) of the server, the O.S. and the DMA driver itself.

Fig. 11 depicts the fluctuations of interconnection throughput when a switch from normal operation OCS to OPS, and the other way round, is initiated by the NFV application and performed in the data plane. Protocol overheads are limiting throughput in OCS whereas for OPS the dummy key characters used for packet synchronization are limiting the max throughput, plus the fact that we transmit/switch OPS in a 50% ratio.

Regarding the energy efficiency of the proposed architecture, it is well known that optical network devices deliver higher port radix with fixed power consumption and non-restrictive switching capacity. On the contrary, electrical ones usually offer less ports with limited switching capacity and higher power consumption values, which can vary depending on the switching traffic load. The main point why all-optical switching is more energy efficient than electrical switching, is due to the removing of the optical transceivers, which count for more than 50% of the total power consumption. Optical switching elements used in this experiment show much lower power consumption (some tens of Watts) than regular electrical switches (several hundreds of Watts). More specifically, the 192×192 OCS switch consumes 75 Watt in regular operation, the 1×4 SSS consumes less than 10 Watt while the total power consumption of the 4×4 SOA-based OPS prototype is 50 Watt. OPS's breakdown contributions are: FPGA controller 15 Watt, label processor 20 Watt and SOA-driver 15 Watt. Furthermore, recent experimental and simulation research [21], [22] has shown significant differentiation in terms of energy efficiency between architectures using all-optical switching and in others using conventional electrical switching equipment.

Fig. 12 shows the message exchanges among the VDC planner, ODL and the OF agents. The first two requests are performed to create the initial OPS-based VDC (only the OPS static flow is shown). The second set of messages reconfigures the network to go through the optical splitter (for OCS multicasting), which consists of the OPS flow deletion, the creation of two cross-connections in the Polatis OCS switch and one reconfiguration command for the Tx NIC. All the sets of requests consist of HTTP requests to the controller from the VDC planner & OpenFlow FLOW_MOD and CFLOW_MOD message sequence exchanges between the controller and the different agents. Other secondary OpenFlow messages are omitted in order to improve readability.

Lastly, in regard to performance of the demonstrated VDC and NFV, not only the total (re)configuration times but also the contribution of each individual element were measured. The total OCS/OPS channel configuration time includes: (i) the ODL SDN controller processing time; (ii) control message transmission time (which depends on the actual experiment setup); and (iii) the device reconfiguration time. Specifically, in this experiment, the SDN controller needs around 210 msec to process requests coming from the RESTful API in order to forward the corresponding OF configuration commands to the OF agents of the network devices. It approximately takes a further 200 msec for those commands to reach the OF agents and to be processed there. At last, Polatis OCS switch, OPS switch, WSS and NIC require around 16, 10, 300 and 18 msec

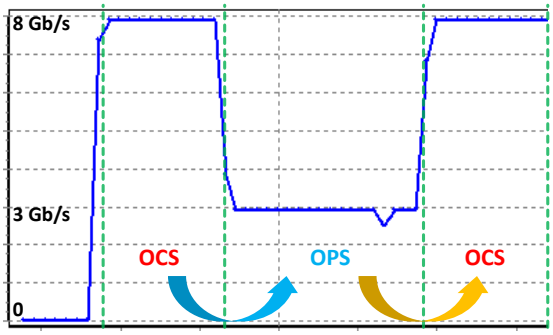

Fig. 11. Throughput plot, illustrating the OCS-to-OPS switch-overs and vice versa

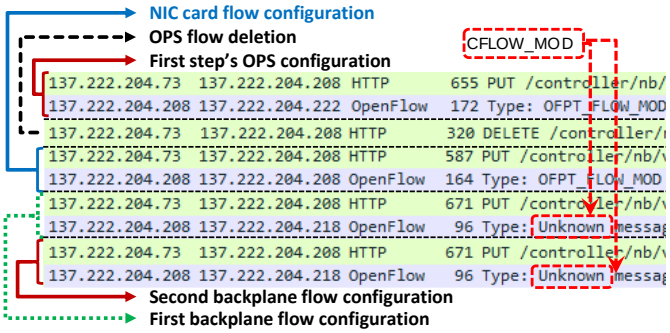

Fig. 12. Message exchange captures from SDN-enabled control plane

respectively to properly configure themselves. The above device configurations of course can be performed in parallel. So, assuming that in order to establish an end-to-end OCS channel we need to successfully configure the optical ToR before configuring NIC, establishing an OCS channel will need 970 msec (also using the WSS) or 690 msec without WSS, while it takes around to 420 msec for OPS connection establishment.

VII. Conclusion

This paper demonstrates for the first time an all-optical programmable DCN architecture enabling OPS/OCS multicasting for realistic monitoring and migration scenarios. The novel networking schemes demonstrated in this paper include a SDN-enabled, virtualize-able and re-configurable optical data plane fully integrated and supported by an extended control plane. In this work, the SDN controller and NFV server are able to provide data plane monitoring and database migration function virtualization, on top of a virtual data center environment realized and administrated by a VDC planner application.

References

[1] G. M. Saridis, S. Peng, Y. Yan, A. Aguado, B. Guo, M. Arslan, C. Jackson, W. Miao, N. Calabretta, F. Agraz, S. Spadaro, G. Bernini, N. Ciulli, G. Zervas, R. Nejabati, and D. Simeonidou, "LIGHTNESS: A Deeply-Programmable SDN-enabled Data Centre Network with OCS/OPS Multicast/Unicast Switch-over," in *European Conference on Optical Communication (ECOC)*, Valencia, 2015, p. PDP 4.2.

[2] "Cisco Global Cloud Index: Forecast and Methodology, 2012–2017," *Cisco*. [Online]. Available: http://cisco.com/c/en/us/solutions/collateral/service-provider/global-cloud-index-gci/Cloud_Index_White_Paper.html. [Accessed: 06-Oct-2014].

[3] P. Kansal and A. Bose, "Bandwidth and Latency Requirements for Smart Transmission Grid Applications," *IEEE Trans. Smart Grid*, vol. 3, no. 3, pp. 1344–1352, Sep. 2012.

[4] H. Ye and Z. Song, "Research on the Next-Generation Green Data Center Technology," in *2013 Fourth World Congress on Software Engineering (WCSE)*, 2013, pp. 257–260.

[5] N. Farrington, G. Porter, S. Radhakrishnan, H. H. Bazzaz, V. Subramanya, Y. Fainman, G. Papen, and A. Vahdat, "Helios: a hybrid electrical/optical switch architecture for modular data centers," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 41, no. 4, pp. 339–350, 2011.

[6] G. Wang, D. G. Andersen, M. Kaminsky, K. Papagiannaki, T. S. Ng, M. Kozuch, and M. Ryan, "c-Through: Part-time optics in data centers," in *ACM SIGCOMM Computer Communication Review*, 2010, vol. 40, pp. 327–338.

[7] Z. Cao, R. Proietti, M. Clements, and S. J. B. Yoo, "Experimental demonstration of dynamic flexible bandwidth optical data center network with all-to-all interconnectivity," in *2014 European Conference on Optical Communication (ECOC)*, 2014, pp. 1–3.

[8] G. Saridis, E. Hugues-Salas, Y. Yan, S. Y. Yan, S. Poole, G. S. Zervas, and D. E. Simeonidou, "DORIOS: Demonstration of an All-Optical Distributed CPU, Memory, Storage Intra DCN Interconnect," in *Optical Fiber Communication Conference*, 2015, p. W1D.2.

[9] G. M. Saridis, Y. Yan, S. Yan, M. Arslan, T. Bradley, N. V. Wheeler, N. H. L. Wong, F. Poletti, M. N. Petrovich, D. J. Richardson, S. Poole, G. Zervas, and D. Simeonidou, "EVROS: All-Optical Programmable Disaggregated Data Centre Interconnect utilizing Hollow-Core Bandgap Fibre," in *European Conference on Optical Communication (ECOC)*, Valencia, 2015, p. Tu 3.6.5.

[10] S. Yan, E. Hugues-Salas, V. J. F. Rancano, Y. Shu, G. Saridis, B. Rahimzadeh Rofoee, Y. Yan, A. Peters, S. Jain, T. May-Smith, P. Petropoulos, D. J. Richardson, G. Zervas, and D. Simeonidou, "Archon: A Function Programmable Optical Interconnect Architecture for Transparent Intra and Inter Data Center SDM/TDM/WDM Networking," *J. Light. Technol.*, vol. 33, no. 7, pp. 1586–1595, 2015.

[11] W. Miao, F. Yan, H. Dorren, and N. Calabretta, "Petabit/s Data Center Network Architecture with Sub-microseconds Latency Based on Fast Optical Switches," in *European Conference on Optical Communication (ECOC)*, Valencia, 2015.

[12] B. A. A. Nunes, M. Mendonca, X.-N. Nguyen, K. Obraczka, and T. Turletti, "A Survey of Software-Defined Networking: Past, Present, and Future of Programmable Networks," *IEEE Commun. Surv. Tutor.*, vol. 16, no. 3, pp. 1617–1634, Third 2014.

[13] A. Lara, A. Kolasani, and B. Ramamurthy, "Network Innovation using OpenFlow: A Survey," *IEEE Commun. Surv. Tutor.*, vol. 16, no. 1, pp. 493–512, First 2014.

[14] M. F. Bari, R. Boutaba, R. Esteves, L. Z. Granville, M. Podlesny, M. G. Rabbani, Q. Zhang, and M. F. Zhani, "Data Center Network Virtualization: A Survey," *IEEE Commun. Surv. Tutor.*, vol. 15, no. 2, pp. 909–928, 2013.

[15] Y. Yan, Y. Shu, G. M. Saridis, B. R. Rofoee, G. Zervas, and D. Simeonidou, "FPGA-based Optical Programmable Switch and Interface Card for Disaggregated OPS/OCS Data Centre Networks," in *European Conference on Optical Communication (ECOC)*, Valencia, 2015.

[16] W. Miao, F. Agraz, S. Peng, S. Spadaro, G. Bernini, J. Perello, G. Zervas, R. Nejabati, N. Ciulli, D. Simeonidou, H. Dorren, and N. Calabretta, "SDN-enabled OPS with QoS guarantee for reconfigurable virtual data center networks," *IEEEOSA J. Opt. Commun. Netw.*, vol. 7, no. 7, pp. 634–643, Jul. 2015.

[17] B. Guo, S. Peng, C. Jackson, Y. Yan, Y. Shu, W. Miao, H. Dorren, N. Calabretta, F. Agraz, J. Perello, S. Spadaro, G. Bernini, R. Monno, N. Ciulli, R. Nejabati, G. Zervas, and D. Simeonidou, "SDN-enabled programmable optical packet/circuit switched intra data centre network," in *Optical Fiber Communications Conference and Exhibition (OFC), 2015*, 2015, pp. 1–3.

[18] "ETSI GS NFV 002 V1.2.1 (2014-12)," 2014. [Online]. Available: http://upcommons.upc.edu/handle/2117/21093.

[19] W. Miao, X. Yin, J. Bauwelinck, H. Dorren, and N. Calabretta, "Performance assessment of optical packet switching system with burst-mode receivers for intra-data center networks," in *2014 European Conference on Optical Communication (ECOC)*, 2014, pp. 1–3.

[20] Broadcom, "Broadcom BCM56850 StrataXGS Trident II Switching Technology." [Online]. Available: http://www.broadcom.nl/collateral/pb/56850-PB03-R.pdf.

[21] M. Imran, P. Landais, M. Collier, and K. Katrinis, "A data center network featuring low latency and energy efficiency based on all optical core interconnect," in *2015 17th International Conference on Transparent Optical Networks (ICTON)*, 2015, pp. 1–4.

[22] Y. Ji, J. Zhang, Y. Zhao, H. Li, Q. Yang, C. Ge, Q. Xiong, D. Xue, J. Yu, and S. Qiu, "All Optical Switching Networks With Energy-Efficient Technologies From Components Level to Network Level," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 8, pp. 1600–1614, Aug. 2014.